Eta_Earth from Kepler 2011 Data
A few points of view

M. Shao

ExoPAG Discussion 6/2/11

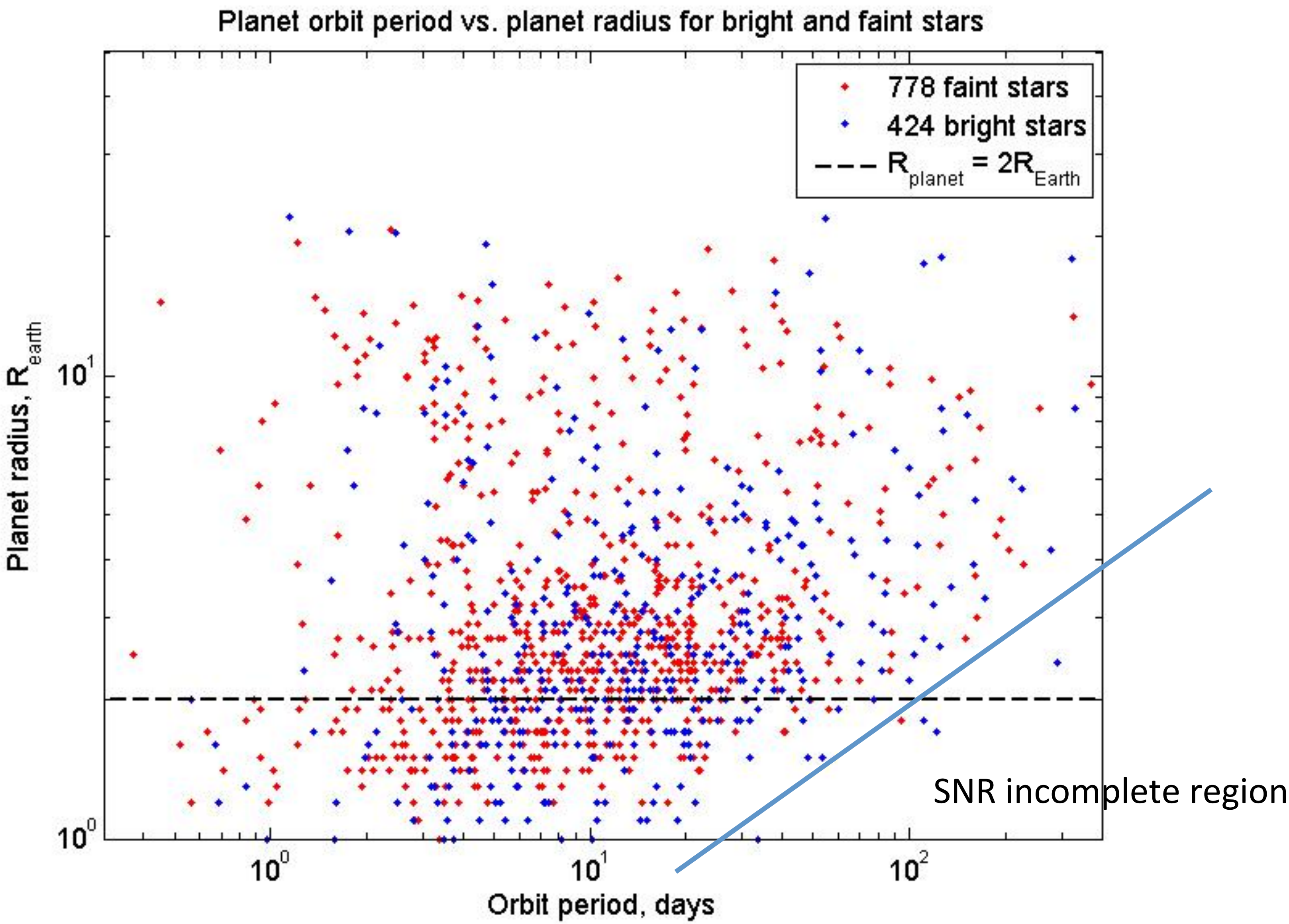
Outline

- The Kepler Planet (candidates)
- Three Papers (submitted) where there's overlap,
  - A. Howard et. al (no estimate of eta_Earth)
  - Catanzarite, Shao (eta_Earth ~ 3%)
  - Traub (eta_Earth ~50%)
- Differences in assumptions
- Consequences of the conclusion
  - Is the Kepler data processing pipeline seriously flawed?
  - How many more planets will/will not be found?
  - Consequences for future Space Missions

The 1235 Kepler Planet (candidates) (from Borucki, feb 2011, posted on AstroPh)

- Planets identified because of **"great improvements to the data-processing pipeline",** page 2
- **"**a total of 1235 planetary candidates that **show transit-like signatures in the first 132 days** of data**",** page 2
- Data on each planet candidate included (Period (days), Radius (R_earth), SNR, plus much more) , page 59
- "SNR is Total **SNR of all transits detected**. SNR=Depth/(Std*sqrt(N)) where Std is the standard deviation of all data outside of transits (**Q0 through Q5**) and N is the total number of measurements inside of all transits." page 59
  - SNR is the **SNR for ~400 days** of Kepler data
- Detection threshold is **SNR >7** to be included in the list of 1235.
- The 1235 planets orbit 997 stars. When single transit planets are removed there are **1202 planets around 979 stars.** Also removed planets > $2R_{jup}$.
- Other papers on Kepler data estimates ~10% false positives on bright stars and <30% false positives on faint stars.

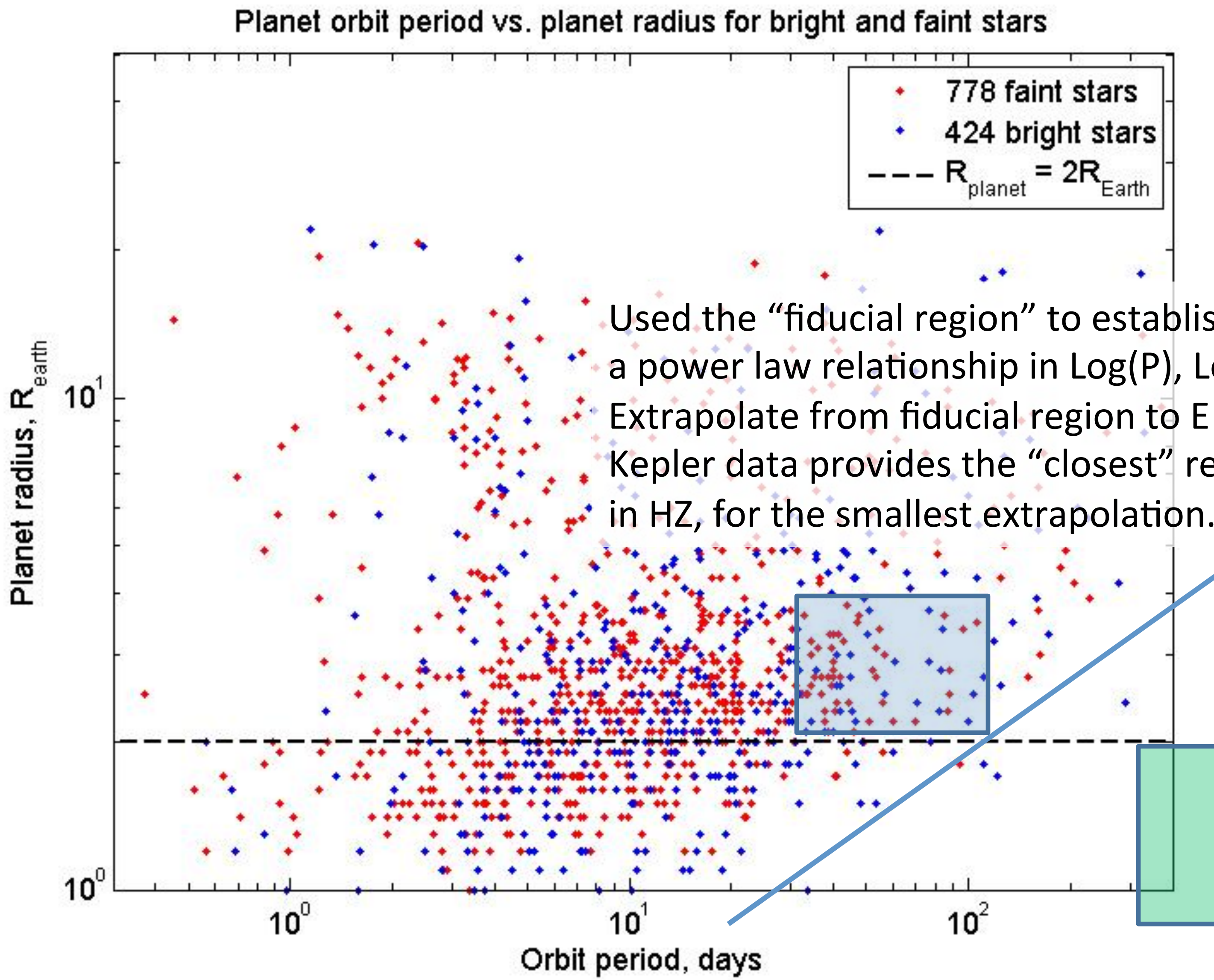
1202 Planet Candidates



Completeness Corrections

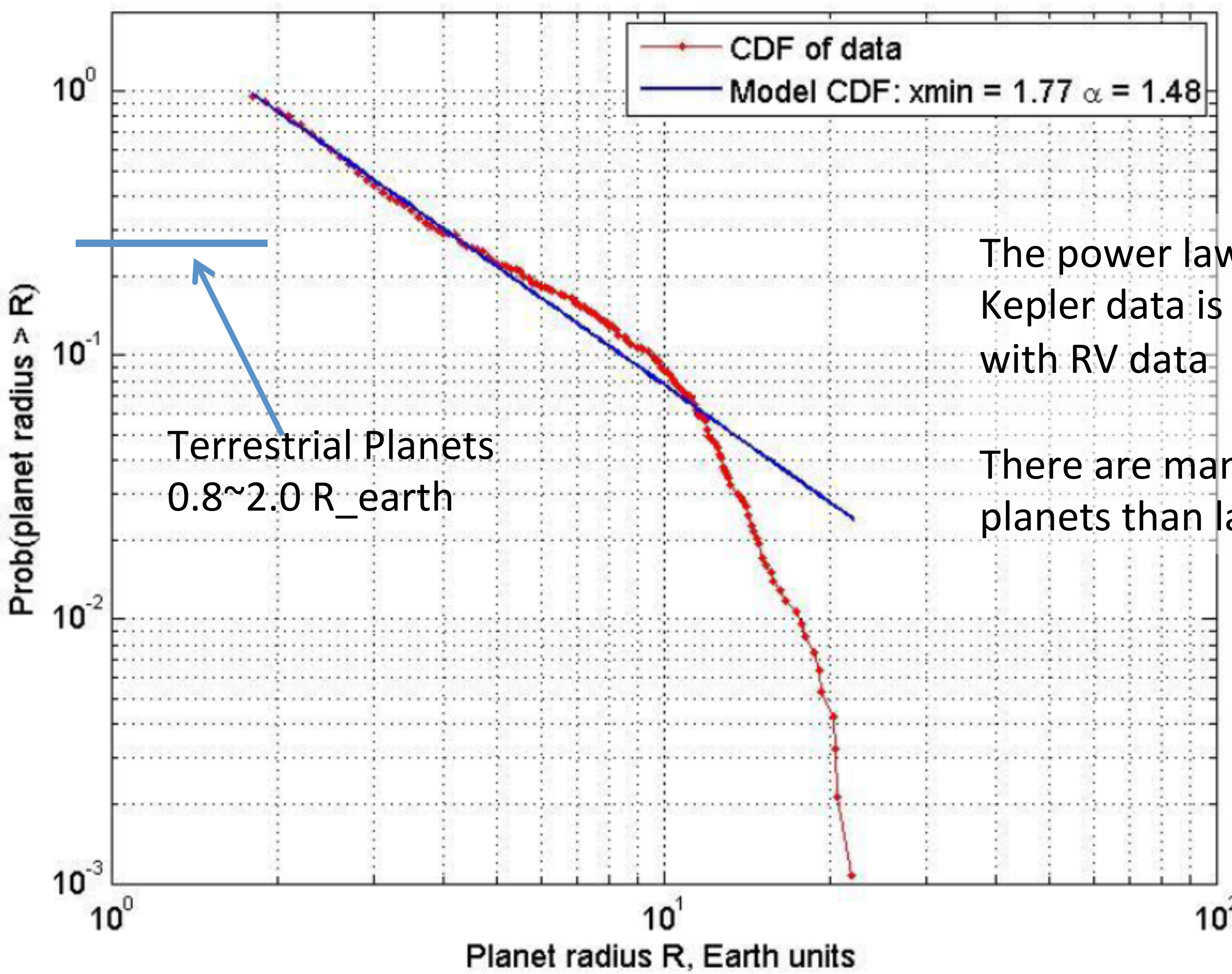
- Orbit inclination completeness (only some orbit inclinations result in a transit). Well known effect. Can be accurately corrected. (R_{star}/Semimajoraxis)
- Timing completeness. The 1202 planets have at least 1 transit during the 1st 132 days, AND at least 2 transits during 1st 400 days. This effect is also well understood and corrected accurately.
  - All planets with periods < 132 days should be detected (If adequate SNR) ~50% of planets with periods ~264 days should be detected.
- SNR completeness. (from discussions/email with N. Gautier after his talk at JPL)
  - Before launch, **Phot_noise= 20ppm** = sqrt(stellar_noise^2+phot_noise^2) stellar_noise = 10ppm, phot_noise(12mag) = 17ppm. (6.5hr transit)
  - Currently **Total_noise 30ppm**, Stellar_noise ~25 ppm, Phot_noise=17ppm
  - From the above one can calculate the SNR of a 2R_earth planet around a G star for a **16 mag** star for **3 transits** in a **120 day** orbit in a 400 day data set to be **SNR~7**.

Extrapolating from Existing data to Earths in HZ



Used the "fiducial region" to establish
a power law relationship in Log(P), Log(Rp)
Extrapolate from fiducial region to E inHZ
Kepler data provides the "closest" region to E
in HZ, for the smallest extrapolation.

Power Law in Planet Radius

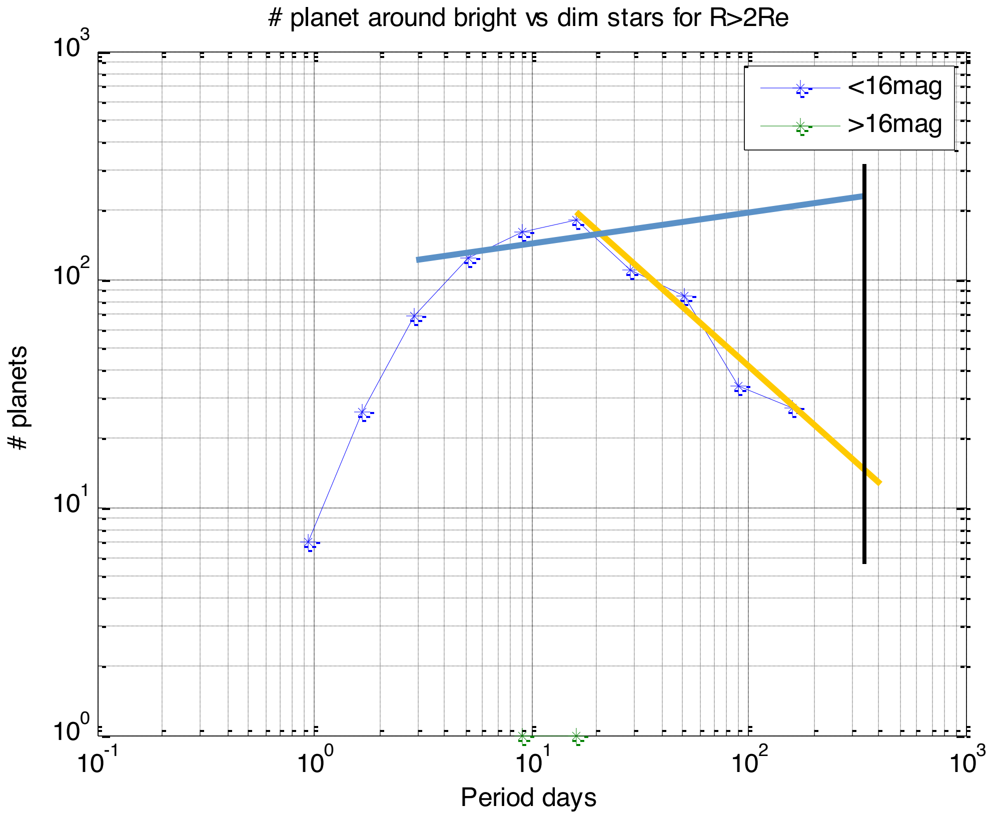


The power law index for Kepler data is consistent with RV data

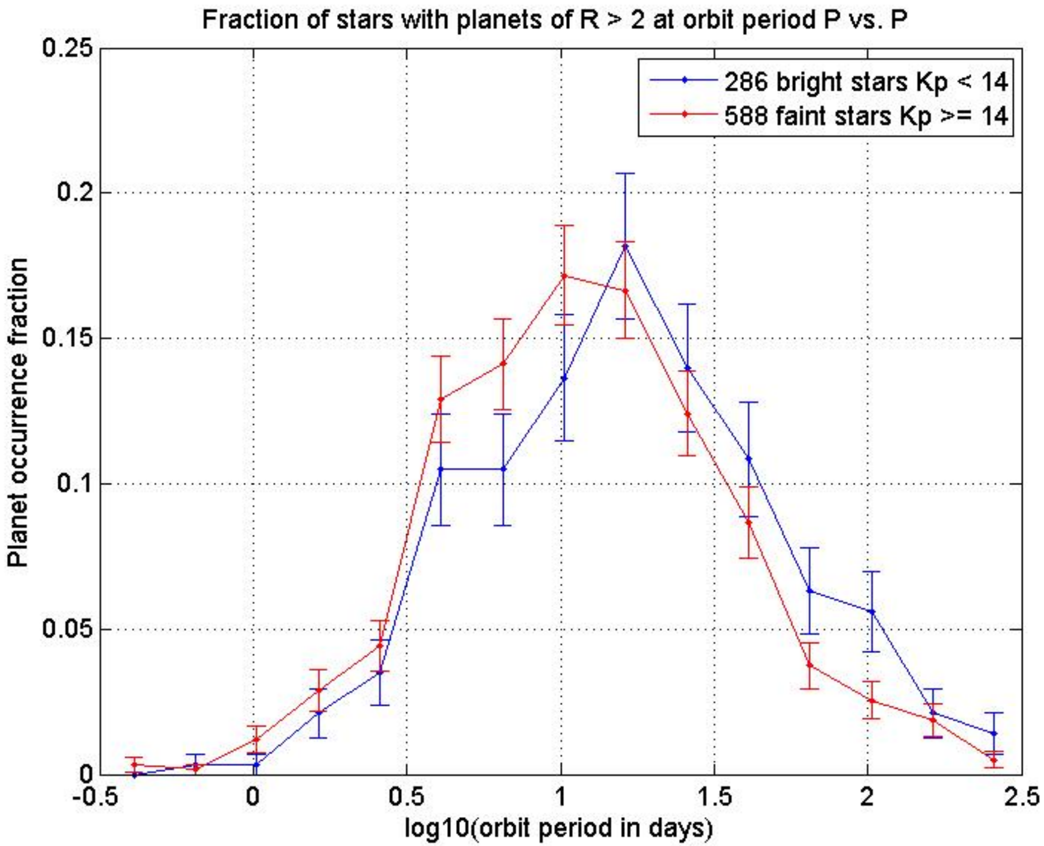
There are many more small planets than large planets.

Figure 3 Planet radius power-law model vs. cumulative distribution function for 1176 planet candidates with $P < 132$ days. The fitted power-law index is $\alpha = 1.48$. The planet radius power-law fit has a lower cutoff of $r_{min} = 1.77$; this is due to incompleteness when the planet is too small to be detected with sufficient SNR. The fit is excellent out to $r \sim 4$.

Power Law with Period

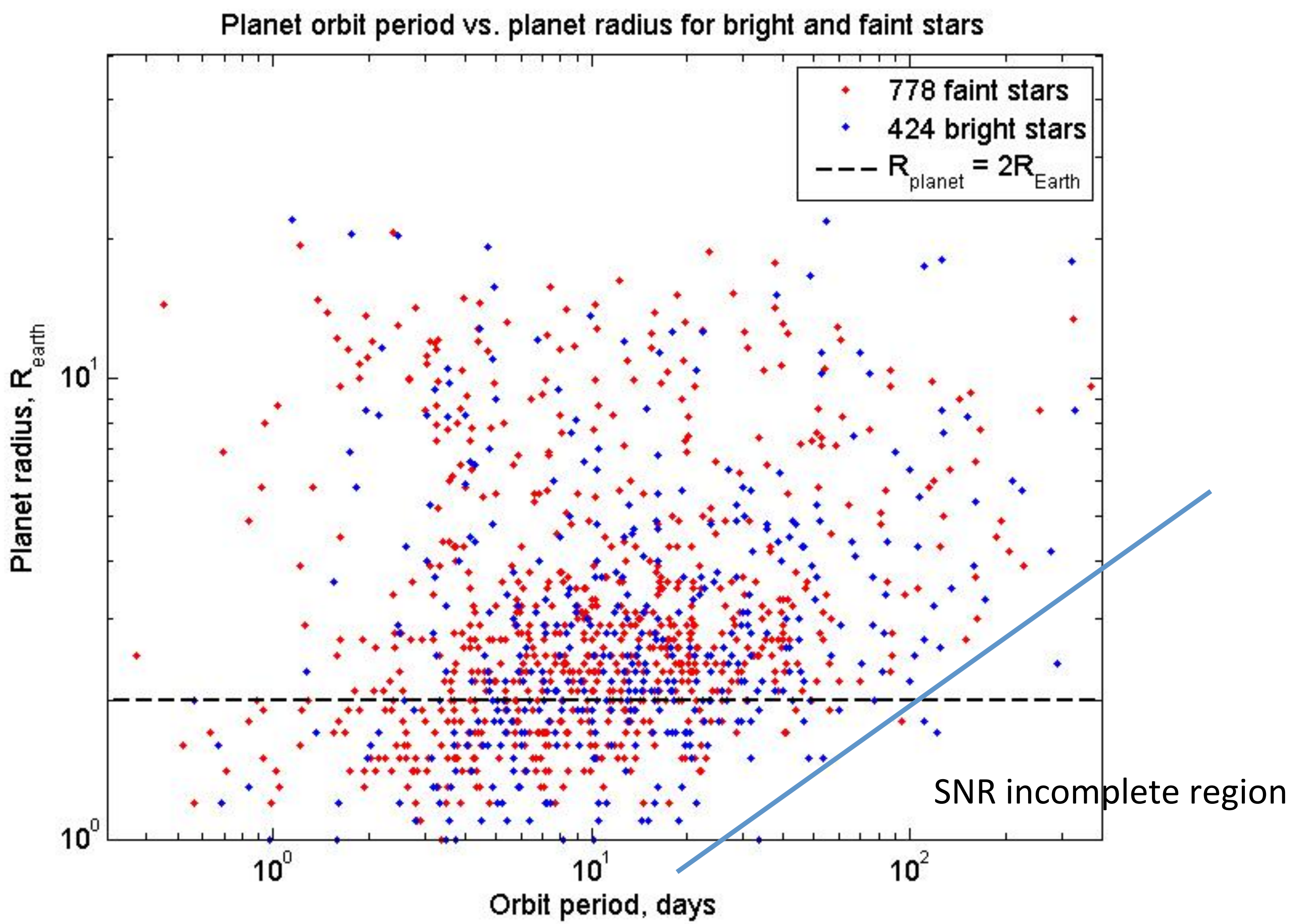


Is there a diff between Bright and Dim Stars?



Fraction of stars with planets larger than 2R_earth as a function of orbit period are similar for bright and faint stars. Error bars show Poisson noise. If the sample were incomplete for periods between 40 and 120 days, we'd expect the planet occurrence rate to drop significantly for faint stars compared to bright stars in that period range. But this is not the case: a Kolmogorov-Smirnov test shows that for periods longer than 20 days and radii larger than 2R_earth, planets orbiting bright stars and faint stars are drawn from the same period distribution with 94% likelihood.

1202 Planet Candidates



Power Law in Semimajor Axis

Cumulative Distribution

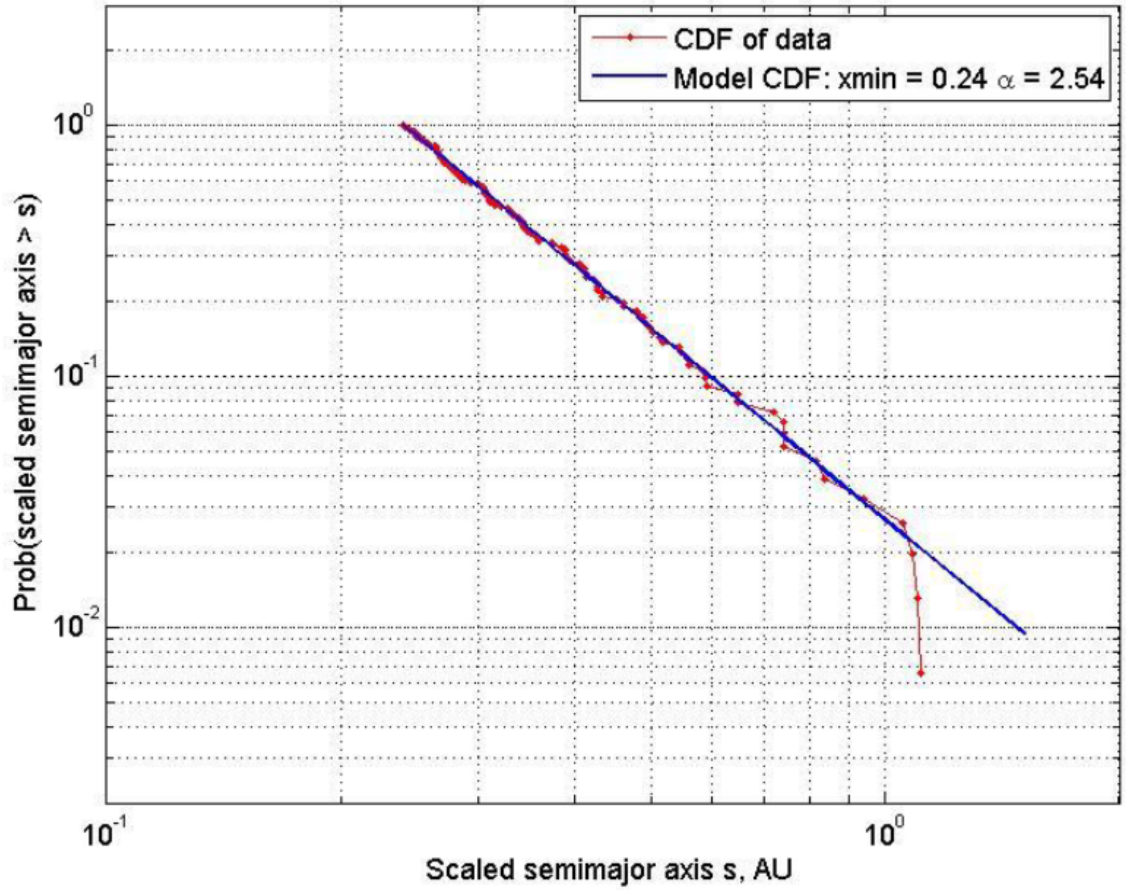


Figure 4 Scaled semimajor axis power-law model vs. cumulative distribution function for 577 Kepler planet candidates with $r < 4$. The fitted power-law index is $\alpha = 2.54$. The scaled semimajor axis power-law fit has a lower cutoff of $s_{min} = 0.24\ AU$, corresponding to an orbit period of about 40 days for a G2 star. The data falls away from the model for $s > 1\ AU$ due to incompleteness in long-period candidates with at least two observed transits.

Data for Planets between
2~4 Rearth
0.24 and 0.5 AU

HZ from 0.75 to 1.8AU
From 0.8~2.0 Rearth

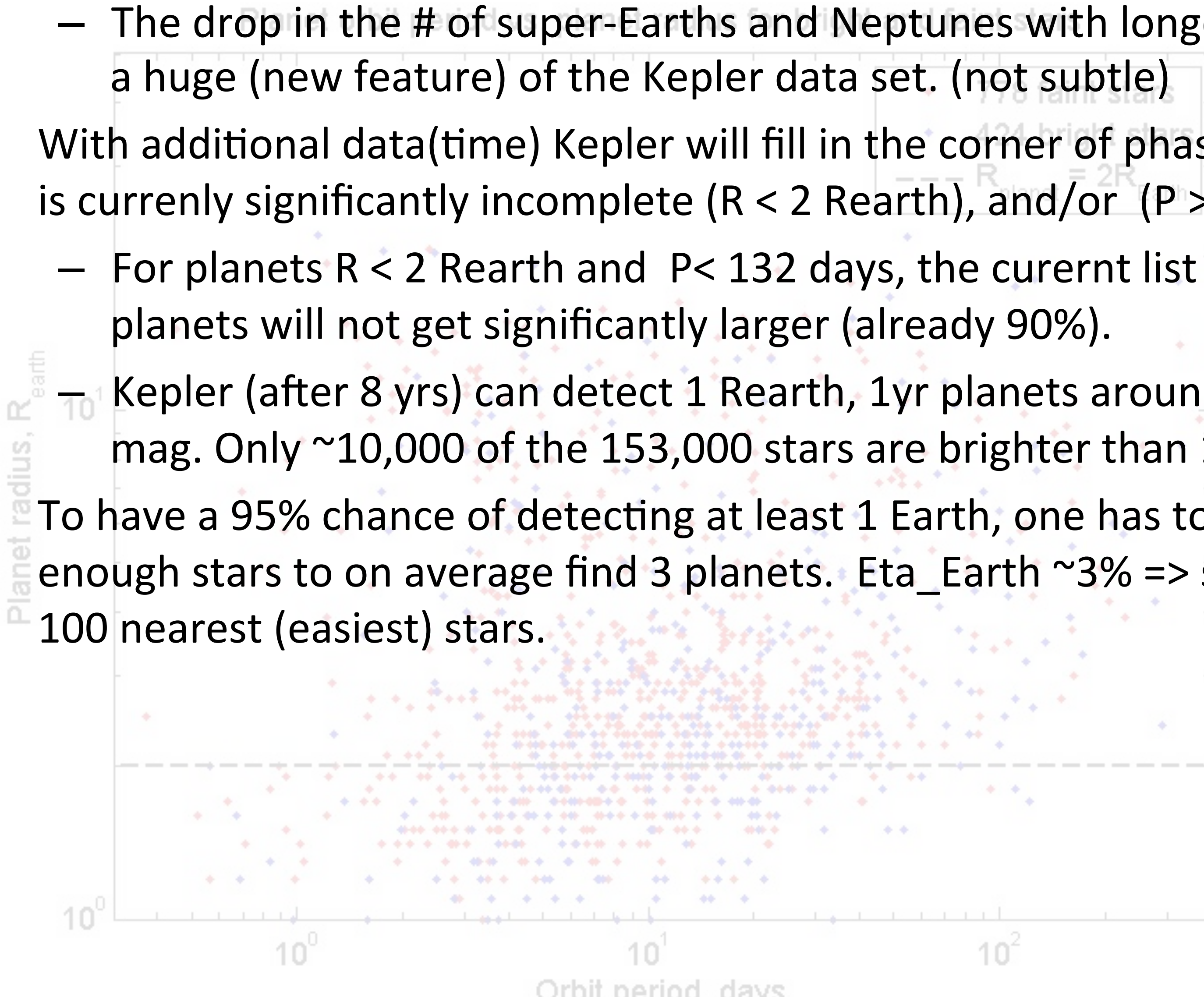
Eta_Earth ~2.9% (1.9%~4.7)
after correcting for orbit inclination.

Different Samples

- Howard paper did not attempt to extrapolate to the HZ. (Howard vs Catanzarite)
  - SNR > 10 in 90days vs SNR>7 in 400 days
  - Star temp between 4100~6100K, vs all stars
  - P<50 days vs P<132 days
  - The two biggest differences in the sample are P<50 days, (misses the break in the power law), SNR>10 much more selective (very low false alarm) but much more incomplete.
- Completeness of longer period planets (30~132 day)
  - For small planets (< 2 Rearth) there is significant incompleteness for periods >30 days. Not the case for medium sized planets (2~4 Rearth), complete to 132 days (and 50% temporally complete to 260 days).

Conclusions

- The vast majority of planets have relatively short periods (3~30 days)
  - The drop in the # of super-Earths and Neptunes with longer periods is a huge (new feature) of the Kepler data set. (not subtle)
- With additional data(time) Kepler will fill in the corner of phase space that is currenly significantly incomplete (R < 2 Rearth), and/or (P > 132 days)
  - For planets R < 2 Rearth and P< 132 days, the curernt list of 1235 planets will not get significantly larger (already 90%).
  - Kepler (after 8 yrs) can detect 1 Rearth, 1yr planets around stars < 12 mag. Only ~10,000 of the 153,000 stars are brighter than 12 mag.
- To have a 95% chance of detecting at least 1 Earth, one has to search enough stars to on average find 3 planets. Eta_Earth ~3% => search the 100 nearest (easiest) stars.



Backup 1 SNR Threshold

- In Howard et. al. 2011, the paper only considered planets with SNR>10 in 90 days of data. The Kepler project considers a transit detection anything with SNR>7 in ~400 days of data. The Higher SNR results in a highly speculative determination of the planet distribution function.
  - High SNR means high confidence the detection is valid. (low false alarm probability) but ALSO, many real planets in the data are now ignored. The data is then replace by an estimate of the SNR completeness.
  - With a lower SNR threshold, the false alarm rate due to SNR is still very low. But instead of an estimate of SNR completeness, we use instead the data of discovered planets (100's of planets).
- A conservative approach to claiming a planet has been detected would favor a high SNR threshold, but this results in a speculative answer deriving the underlying planet distribution function. A Conservative approach for estimating the Planet dist function is to take a lower SNR.
  - Kepler looks at 150,000 stars, and over 1 year has ~1e8 photometric measurements. As long as the SNR is high enough that false alarms due to SNR glitches is << the actual number of planets found, we